Citation Evidence Report

EB-2 NIW Petition — National Interest Waiver

Matter of Dhanasar · Prong 2 (well-positioned)

David Botstein

Princeton University

Google Scholar profile

Generated 2026-05-21 by CiteMap. This report organises Google Scholar citation data into the structure USCIS adjudicators apply to Prong 2 of Matter of Dhanasar (the petitioner is well positioned to advance the proposed endeavor) — the prong where past citation evidence is most probative. It is a drafting aid for the petitioner's counsel — not legal advice, and not a guarantee of any outcome. All figures must be verified, and citation counts re-snapshotted as of the petition filing date, before use in a filing.

A. Overview & Filtering Statement

14	14	3	194
Citing papers mapped	Citation edges	Home papers mapped	h-index (GS)

Filtering statement — methodology & limits

Citation **independence** is classified per citing paper by comparing the citing paper's authors to this scholar. *Self* citations are those where the scholar is an author of the citing work; *co-author* citations are by the scholar's known collaborators; *same-institution* citations are by authors affiliated with the scholar's institution(s); all remaining classified citations are *independent*. Per AAO practice, only independent citations are treated as probative of influence beyond the scholar's own circle.

Known limitations — counsel must verify. (1) Collaborator identification draws on the co-author list published on the Google Scholar profile; a collaborator not listed there may be missed, so the independent share below should be read as an **upper bound**. (2) Citation counts are a crawl-time snapshot; eligibility is judged as of the petition filing date and post-filing citations carry no weight — re-snapshot before filing. (3) Citations that could not be classified (no author data) are excluded from the percentages and reported separately.

B. Citation Independence

The AAO credits citations only where they show influence **beyond the scholar's own circle**. Self-citations and co-author citations are expressly discounted; the independent share below is the load-bearing figure.

100.0% independent of 12 classified citing papers

Citation type	Count
Independent	12
Self-citation	0
Co-author	0
Same-institution	0

2 citing papers could not be classified (no author data) and are excluded from the percentages above.

C. Significant Contributions & Their Citation Evidence

Each contribution below is presented as the AAO expects: a specific claim, followed by the **independent** citation evidence for the paper(s) that carry it. Citation counts are stated **per article**, never as a body-of-work total — the AAO holds aggregate totals to be a final-merits signal, not Criterion-5 evidence.

Where the data allows, a paper also shows its **field-normalised** standing — how its citation count ranks against Semantic Scholar papers in the same field and publication year. The comparison field is named explicitly; counsel should confirm it is the appropriate one, as the AAO scrutinises a petitioner's choice of comparison field.

Contribution 1

Claim — Contribution 1

The researcher developed foundational methods for clustering and visualizing genome-wide expression patterns, establishing a standard approach for analyzing high-dimensional biological data.

The researcher’s primary contribution rests on the 1998 paper 'Cluster analysis and display of genome-wide expression patterns.' This work appears to have introduced a systematic framework for organizing and interpreting complex genomic datasets, addressing the challenge of making sense of large-scale expression data. By focusing on both analysis and display, the titles suggest a dual emphasis on computational rigor and intuitive visualization, which was likely novel at the time of publication.

The significance of this line of work is evidenced by its substantial citation count of over 21,000, indicating widespread adoption across the scientific community. Furthermore, analysis of citing papers reveals that 100% of the classified citations originate from independent researchers, rather than the author’s own network. This high degree of independent uptake suggests that the methodology has become a standard tool in the field, utilized broadly by diverse groups to advance their own research objectives.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 5

CORE PAPER

Cluster analysis and display of genome-wide expression patterns

1998 · 21,073 citations (GS)

Field-normalised: 17,168 Semantic Scholar citations place it in the top 1% of Biology papers from 1998 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	Using MetaboAnalyst 4.0 for Comprehensive and Integrative Metabolomics Data Analysis (2019)	McGill University, University of Alberta	Canada	—
2	Complex heatmap visualization	National Center for Tumor Diseases	Germany	—
3	Single-cell RNA sequencing technologies and bioinformatics pipelines (2018)	Kyung Hee University, Yonsei University	South Korea	**Methodology**
4	Tumor-associated macrophages restrict CD8+ T cell function through collagen deposition and metabolic reprogramming of the breast cancer microenvironment (2024)	Institute for Research in Biomedicine (IRB), Università della Svizzera italiana, University of California, Los Angeles, University of California San Francisco	Switzerland, United States	—
5	A high-bias, low-variance introduction to Machine Learning for physicists	Boston University, The Graduate Center, City University of New York, University of California	United States	Background

Independent citing papers only; self- and co-author citations excluded. The **S2** column carries Semantic Scholar’s read of each citation — *Methodology / Result* (the citing work used the method or built on the finding — the “built on / relied upon” pattern the AAO credits), *Influential* (S2’s isInfluential signal, Valenzuela et al. 2015), or *Background* (a passing mention).

Citing-text excerpts — how the field used this work

> **METHODOLOGY** Single-cell RNA sequencing technologies and bioinformatics pipelines
>
> *“These methods can be categorized into machine learning-based 73 – 75 , co-expression-based 76 , model-based 77,78 , and information theory-based approaches.”*

Contribution 2

Claim — Contribution 2

The researcher developed the Gene Ontology framework, a seminal tool for the unification of biological data that has become a foundational standard in the field.

CLAIM: The researcher’s primary contribution is the development of the Gene Ontology, introduced in the 2000 Nature Genetics paper titled 'Gene ontology: tool for the unification of biology.' This work stands as a singular, foundational achievement without subsequent follow-up papers by the same author in this specific line of inquiry.

ORIGINALITY: The title suggests the work addressed a critical need for standardization and integration within biological research. By proposing a tool for the 'unification of biology,' the researcher appears to have created a common language or framework to organize disparate biological data, addressing fragmentation in the field at the turn of the millennium.

SIGNIFICANCE: The impact of this contribution is evidenced by its extensive citation record, with over 51,000 citations indicating widespread adoption. Furthermore, analysis of citing papers reveals that 100% of the classified citations originate from independent researchers, demonstrating that the work has been embraced and utilized by the broader scientific community rather than just the researcher's immediate circle.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 2

CORE PAPER

Gene ontology: tool for the unification of biology

2000 · Nature Genetics · 51,366 citations (GS)

Field-normalised: 41,835 Semantic Scholar citations place it in the top 1% of Biology papers from 2000 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	edgeR v4: powerful differential analysis of sequencing data with expanded functionality and improved support for small counts and larger datasets	Genentech Inc, WEHI	Australia, United States	—
2	SRplot: A free online platform for data visualization and graphing (2023)	Shanghai NewCore Biotechnology, Shanghai NewCore Biotechnology Co., Ltd., Shenzhen Ping'an Financial Technology Consulting Co. Ltd	China	**Methodology**

Independent citing papers only; self- and co-author citations excluded. The **S2** column carries Semantic Scholar’s read of each citation — *Methodology / Result* (the citing work used the method or built on the finding — the “built on / relied upon” pattern the AAO credits), *Influential* (S2’s isInfluential signal, Valenzuela et al. 2015), or *Background* (a passing mention).

Citing-text excerpts — how the field used this work

> **METHODOLOGY** SRplot: A free online platform for data visualization and graphing
>
> *“In addition, SRplot contains some commonly used databases on the server, such as genomes/transcriptomes/GO [6] /KEGG [7] databases from human/mouse/rat, and the TCGA data (in MAF format).”*

Contribution 3

Claim — Contribution 3

> *The researcher established a foundational molecular classification framework for human breast tumors, fundamentally shifting the understanding of breast cancer heterogeneity through high-impact publication in Nature.*

CLAIM: The researcher's seminal contribution is the establishment of a molecular classification system for human breast tumors, anchored by the 2000 Nature paper 'Molecular portraits of human breast tumours.' This work serves as the core reference point for this line of inquiry, with no subsequent follow-up papers by the researcher listed in the provided data.

ORIGINALITY: The title suggests a departure from traditional histological classification by proposing distinct molecular profiles for breast tumors. This approach appears to address the need for a more nuanced understanding of tumor biology, potentially revealing underlying genetic or expression-based subtypes that were not previously characterized in such detail.

SIGNIFICANCE: The core paper has accumulated 23,129 citations, indicating it is a highly influential and widely recognized reference in the field. Furthermore, analysis of 12 citing papers reveals that 100% are from independent researchers, demonstrating that the work has been adopted and built upon by the broader scientific community rather than just the researcher's immediate circle.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 5

CORE PAPER

Molecular portraits of human breast tumours

2000 · Nature · 23,129 citations (GS)

Field-normalised: 16,092 Semantic Scholar citations place it in the top 1% of Medicine papers from 2000 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	Advances in systemic therapies for triple negative breast cancer (2023)	Mayo Clinic	United States	—
2	Deciphering breast cancer: from biology to the clinic (2023)	The Walter and Eliza Hall Institute of Medical Research, University of Auckland	Australia, New Zealand	—
3	Breast Cancer—Epidemiology, Risk Factors, Classification, Prognostic Markers, and Current Treatment Strategies—An Updated Review (2021)	Center of Oncology of the Lublin Region St. Jana z Dukli, Medical University of Lublin	Poland	Background
4	Breast cancer: pathogenesis and treatments	Fudan University, Guiyang Maternal and Child Health Care Hospital & Guiyang Children's Hospital	China, P. R. China	—
5	Towards targeting the breast cancer immune microenvironment (2024)	Peter MacCallum Cancer Centre, The University of Melbourne, ZAS Ziekenhuizen	Australia, Belgium	—

Independent citing papers only; self- and co-author citations excluded. The **S2** column carries Semantic Scholar's read of each citation — *Methodology / Result* (the citing work used the method or built on the finding — the "built on / relied upon" pattern the AAO credits), *Influential* (S2's isInfluential signal, Valenzuela et al. 2015), or *Background* (a passing mention).

D. Citing-Institution Prestige & Geography

Top citing institutions

Institution	Country	World ranking	Citing papers
Yonsei University	South Korea	SCImago #238 · THE 86 · QS 50	1
McGill University	Canada	SCImago #168 · THE =41 · QS 27	1
The Graduate Center, City University of New York	United States	—	1
University of Auckland	New Zealand	SCImago #618 · THE =156 · QS 65	1
Unlearn.AI	United States	—	1
Peter MacCallum Cancer Centre	Australia	SCImago #877	1
The Walter and Eliza Hall Institute of Medical Research	Australia	SCImago #580	1
University of California San Francisco	United States	SCImago #98	1
University of California, Berkeley	United States	SCImago #95 · THE 9 · QS =17	1
Kyung Hee University	South Korea	SCImago #792 · THE 251–300 · QS =331	1
Boston University	United States	SCImago #272 · THE =76 · QS =88	1
Mayo Clinic	United States	SCImago #88	1
University of California, San Francisco	United States	SCImago #98	1
University of Michigan	United States	SCImago #43 · THE 23 · QS 45	1
University of Alberta	Canada	SCImago #262 · THE 119 · QS =94	1

Geographic distribution of citing authors

Country	Citing papers
United States	4
Australia	3
China	2
Germany	1
New Zealand	1
Poland	1
P. R. China	1
South Korea	1
Switzerland	1
Canada	1
Belgium	1

Citing-institution prestige and the spread of citing countries speak to recognition **beyond the scholar's own institution and circle** — the dispersion the AAO looks for. World rankings (SCImago / THE / QS) are context, not a stand-alone criterion: the AAO does not treat a citing institution's rank as probative on its own.

E. Citation Growth Over Time

Distinct citing papers by publication year. Sustained or rising citation activity supports continuing relevance; note that only citations **as of the filing date** are weighed by USCIS.

F. AAO Precedent Considerations

Pre-filing self-check (AAO denial patterns)

The AAO non-precedent decisions reject citation evidence on a small set of recurring grounds. Confirm the petition addresses each before filing:

- ☐ Self-citations are disclosed and netted out — a Google Scholar total alone is faulted (§1.1).
- ☐ Evidence is per individual article, not a body-of-work aggregate total (§1.2).
- ☐ The petition articulates why the citations show major significance — numbers never stand alone (§1.5).
- ☐ For the strongest papers, citation content shows the work was built on / relied upon, not just listed (§1.6, §2.2).
- ☐ Co-author / collaborator citations are identified and not counted as independent (§1.7).
- ☐ Recognition is shown beyond the scholar's own institution and circle (§1.8).
- ☐ Every citation figure is snapshotted as of the filing date; post-filing citations are excluded (§1.9).
- ☐ Journal impact factor / downloads are not relied on as proxies for article significance (§1.10, §1.12).
- ☐ For large-collaboration papers, the scholar's specific role is documented (§1.13).
- ☐ Aggregate totals / h-index / field-relative rates are placed in a clearly-labelled final-merits section, per Kazarian (§3, §6.1.7).

Disclaimer

The AAO decisions referenced here are **non-precedent** — persuasive illustrations of how USCIS reasons, not binding law. This report is a drafting aid produced from public citation data; it is not legal advice and does not assess the petition's merits. All analysis must be reviewed by qualified immigration counsel.

G. Citation Evidence Index

Cross-reference of each contribution to the regulatory criterion it supports. Counsel should map these to the petition's exhibit numbers.

Contribution	Core paper	Indep. cites	Supports
Contribution 1	Cluster analysis and display of genome-wide expression patterns	5	Dhanasar — Prong 2 (well-positioned)
Contribution 2	Gene ontology: tool for the unification of biology	2	Dhanasar — Prong 2 (well-positioned)
Contribution 3	Molecular portraits of human breast tumours	5	Dhanasar — Prong 2 (well-positioned)